DOCUMENT RESUME

ED 402 330 TM 025 832

AUTHOR Phelps, Richard P.
TITLE Test Basher Benefit-Cost Analysis.
PUB DATE Mar 96
NOTE 18p.; Version of a paper presented at the Annual Meeting of the American Education Finance Association (1994).
PUB TYPE Journal Articles (080) -- Speeches/Conference Papers (150)
JOURNAL CIT Network News & Views; v15 n3 p1-16 Mar 1996

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *Cost Effectiveness; Educational History; Elementary Secondary Education; Equal Education; Public Opinion; *Standardized Tests; *State Programs; Test Bias; *Testing Programs; *Test Use; Test Validity

ABSTRACT

Starting in the late 1980s, two teams of researchers, well known for their criticism of standardized tests on equity and validity grounds, began attacking standardized testing on efficiency grounds as well, using cost-benefit analysis to do it. Their analyses are reviewed, and their conclusions discussed. The first team, Lorrie A. Shepard, Amelia E. Kreitzer, and M. Elizabeth Graue, wrote "A Case Study of the Texas Teacher Test," published as a report of the Center for Research on Evaluation, Standards, and Student Testing and as an article in "Education Researcher." The second team, Walter M. Haney, George F. Madau, and Robert Lyons, wrote the book "The Fractured Marketplace for Standardized Testing," a survey of the industry and a cost-benefit analysis. Correcting for some of the more obvious mistakes of the authors of the analysis of the Texas program pushes the program's net benefits into the black by a wide margin. The analysis of Haney et al. also fails to convince that the costs outweigh the benefits. Concerned and well-trained educators in state agencies and local school districts make the decisions to purchase or develop standardized tests for reasons that are neither unfair nor sinister. They believe that the benefits outweigh the costs, and this belief is shared by much of the public. (Contains 7 tables and 16 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 402 330

NETWORK NEWS & VIEWS

Vol. XV, No. 3 March 1996

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
RICHARD P. PHELPS
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

EDUCATIONAL EXCELLENCE NETWORK
A Project of **Hudson Institute**

TM025832

BEST COPY AVAILABLE

N E T W O R K

NEWS & VIEWS

A publication of the
Educational Excellence Network

Editorial Office: 18 Lamp Post Court, Potomac, MD 2085-
Subscription Office: 223 N. Guadalupe Street #305, Santa Fe, NM 875-

March 1996

Dear NETWORK Member:

You hold in your hand a terrific issue of *News & Views* . You'll find (in "Chart(er)ing the Course") a major compilation on charter schools, including the Network's own new "First Look" report by Louann Bierlein, Bruno Manno and Chester Finn, as well as news of good schools, legislative victories and setbacks from Massachusetts to New Jersey, from Washington State to Virginia and Florida.

Look, too, at the brimful "Standards and Measures" section, where you'll find previously-unpublished testimony by Pennsylvania Education Secretary Eugene Hickok on why the school establishment ought not dominate the setting of education standards, and an original piece by Richard Phelps on the cost/value of standardized testing.

Other original contributions include Guilford County (N.C.) Superintendent Jerry Weast's congressional testimony on a promising teacher-quality-monitoring system in his community. We've also reprinted the full text of Quentin Quade's excellent essay on school choice, which suffered from production glitches in our January-February issue. For another perceptive--and heterodox--view of school choice, see George Pieler's essay on how government-funded scholarships (aka vouchers) in the District of Columbia may weaken the burgeoning private scholarship movement.

And plenty more for enjoyment and enlightenment. Read on.

Checker Finn Diane Ravitch

PS: Don't forget to check out--and add to--the Network's new Web-site, now on line and worth your time, thanks to the heroic labors of our Web editor, Hudson research assistant Gregg Vanourek. There you'll find timely, useful information and opinion about education developments, an invitation to tell others what's happening in your part of education-land, and a way to pose research questions (and answers) for fellow Network-members. You can find us at **http://www.edexcellence.net**. See you in cyberspace!

The Educational Excellence Network is a project of
Hudson Institute
Herman Kahn Center
P.O. Box 26-919
Indianapolis, Indiana 46226

Test Basher Benefit-Cost Analysis

By Richard P. Phelps

An earlier version of this article was presented at the
1994 Annual Meeting of the American Education Finance Association.
The views expressed here are the author's own and not those
of the American Institutes for Research

Introduction

Within the education research community, mostly in education schools on our college campuses, exists a group of researchers who spend a great deal of their time criticizing standardized testing. For years, their criticisms focussed on the alleged shortcomings of standardized tests with regard to their validity and equity aspects. To be fair, some of the alleged shortcomings were real, and still exist today. For example, off-the-shelf commercially-produced tests from a test publisher with a national market tend to have validity problems, the most obvious of which is that the content of the test may not be highly correlated with the content of the curriculum in any given school where the test is used. Such a test is fine to use if the object is to monitor the school's program or to roughly compare a student's progress to that of a national average. But, using such a test to measure and judge an individual student's performance in a curriculum different from that embodied in the test seems unfair. States and public school districts have responded to such criticisms by making their high-stakes tests — those tests with serious consequences for individual students — curriculum-derived (i.e. "criterion-referenced")

More recently, some of standardized testing's critics have expanded the breadth of their assault to include testing's costs.

It is generally agreed that, as tests go, ordinary standardized student tests have considerable cost advantages. It would take relatively enormous resources, for example, for an individual school to develop tests from scratch that contain the reliability and comparative properties of the standardized tests produced by commercial vendors, state education agencies, and some of the largest school districts.

To people outside the field, then, the cost of standardized student testing would likely seem a rather straightforward, mundane topic. But, within the field, it's an anxiety-producing subject that spawns tense arguments. The arguments tend to turn on the worth or intrinsic educational value of the tests themselves, the amount of time taken up by test-taking and test-preparation, and the assignment or lack thereof of particular cost components as attributable to standardized testing.

If one chooses to believe that standardized test-taking and test-preparation time have no intrinsic instructional value and, further, that standardized tests are separate from and contribute nothing to the instructional plan of a school, then one might well consider standardized tests to be very costly, because they take up time that might otherwise be devoted to instruction. To such critics, the problematic costs associated with standardized tests are not represented by the purchase price paid to the commercial vendors but, rather, by the lost opportunity for learning that could have taken place in the time devoted to taking and preparing for standardized tests.

Starting in the late 1980s, two teams of education researchers, well-known for their criticism of standardized tests on equity and validity grounds, began attacking standardized testing on efficiency grounds as well, using benefit-cost analysis to do it. These two teams included some of the most visible education researchers in the country — highly-regarded, well-positioned, oft-honored, "top" researchers on the subject of educational testing.

I will review the benefit-cost analyses of these two teams of researchers, summarizing their assumptions and analyzing their conclusions. The first team — Lorrie A. Shepard, Amelia E. Kreitzer, and M. Elizabeth Graue — wrote "A Case Study of the Texas Teacher Test," which was published as a CRESST report and as an article in *Education Researcher*. (Shepard, et.al, 1987) The second team — Walter M. Haney, George F. Madaus, and Robert Lyons — wrote the book *The Fractured Marketplace*

for Standardized Testing, which is both a survey of the industry and a benefit-cost analysis of standardized testing in general. (Haney, et.al, 1993)

Methodological Background

Most readers probably appreciate the value of benefit-cost analysis as an analytical construct. Benefit-cost analysis is imbedded in all studies that ask the essential question of an activity, "Is it worth doing?" Benefit-cost analysis is a set of techniques, heuristics, philosophy, and logic that can impose an order and rigor on the process used to answer the essential question.

The logic of benefit-cost analysis is that of the accountant's spreadsheet. Indeed, one could accurately describe it as economists' accounting method. The essential idea is to capture all relevant costs and benefits, broadly considered, on one sheet of paper and weigh them in the balance. If the enterprise or project shows more benefit than cost (i.e. net benefits are positive) it can be said to be economically worthwhile. It is assumed that the researcher will do an honest and responsible job of trying to capture all the relevant benefits and costs. If they can't be estimated with any precision, the researcher should at least enumerate them and leave it to the reader to estimate their value.

What one person considers to be a benefit, however, another person may not. Indeed, what one person considers to be a benefit, another person may regard as a cost. The details of benefit-cost analyses, then, are often subject to debate. It is, however, considered incumbent upon the researcher to properly identify what perspective she is adopting. Ideally, a benefit-cost analysis calculates the benefits and costs as they accrue to all of society — such is the nature of a *social* benefit-cost analysis. Anything less — an analysis that calculates benefits and costs for a sub-group — is a *private* benefit-cost analysis, and the researcher is obligated to explicitly declare it as such.

Benefit-cost analysis should be most welcome in education research. Benefit-cost analysis imposes a structure in which "the whole picture" gets considered. It provides a framework that can impose rigor and honesty onto evaluations that could otherwise be sloppy.[0]

By the same token, most readers are probably also well aware of how benefit-cost analysis can be misused. A researcher can make unreasonable or dishonest estimates, ignore some relevant benefits or costs, include some irrelevant benefits or costs, or double count. There can be a tendency among advocates to exclude or include benefits or costs according to their preferences.

What costs and benefits *are* relevant? Generally, they are the marginal costs or benefits that are attributable to the activity in question and not another activity. When someone argues that the cost of a test is X, the appropriate cost to cite is the marginal cost of the test, the cost that can be attributed to the existence of the test and not to any other activity. Looked at another way, a marginal cost of a test is a cost that is *caused* by the test, one that doesn't exist without the test. An heuristic one can use to determine if an activity is a marginal cost of a test or not: take the test away and see if the activity disappears.

All the cost elements involved in testing are associated with gross costs, but not all produce marginal costs. It is undeniably true, for example, that at least three major cost components are always involved in any standardized student test. Personnel time, student time, and physical overhead are necessary components for the administration of any student test; no student test can be conducted without students, some educator involvement, even if just in monitoring, and a classroom. An exercise in counting the gross costs of testing would count all three. An exercise in counting the marginal costs of testing, however, would not.

Take physical overhead, for example. It goes without saying that student test takers need classroom space in which to write their exams. Should the capital and maintenance costs of that classroom space be counted as a cost of testing? As gross costs, yes; as marginal costs, no. The reason is that the classrooms are already built to house regular classroom work; they weren't built and they

---

[0] The reader new to the subject is invited to review one of the several excellent guidebooks on the topic, such as that by Gramlich. Other, briefer explanations of benefit-cost analysis can be found in chapters of Stokey and Zeckhauser, or Friedman, or, for that matter, in most intermediate-level economics texts or introductory-level public finance texts. Catterall has written an edifying overview of the problems of applying the benefit-cost model to student testing.

aren't maintained because of the test. If the school decided not to test, the classrooms would still need be there.

The Texas Teacher Test

In the late 1980s, Lorrie Shepard, Amelia Kreitzer, and M. Elizabeth Grau attempted a benefit-cost analysis of the Texas Examination of Current Administrators and Teachers (TECAT). Their work was sponsored by the publicly-funded Center for Research on Evaluation, Standards, and Student Testing (CRESST) at UCLA.

The TECAT was a paper-and-pencil test of basic literacy skills. It was born out of a concern on the part of the citizens of Texas and their elected representatives that weakly-regulated state teacher colleges were producing graduates unqualified to teach. Originally, the TECAT was intended to include both a basic literacy section and a content-area section, but the former section alone survived the deliberations among the legislature and the various interested groups in time for the test administration in March of 1986.

The authors portrayed the acceptance of the TECAT as a *quid pro quo* of the teacher's union, agreed to in return for a salary increase in the midst of the state fiscal crisis caused by the collapse in oil prices in the early 1980s. The test had high stakes — a teacher was required to pass to maintain teacher certification. The test was also widely regarded to be extremely simple. The authors reported that state newspapers displayed easy questions from the TECAT alongside stories of large numbers of teachers participating in study sessions for the exam.

The authors had many criticisms of the test itself and the testing program. Included among them was the assertion that the costs were enormous and easily outweighed the benefits. The authors were not shy about stating their opinions, either. Points critical of the TECAT were listed along with the authors' preferred alternative. Even in the executive summary, one finds editorial comments such as:

- "An atmosphere of stress and bitterness was created by the high-stakes, of literally losing your job if you failed. *Many said the effect would have been different if not passing meant having to take a college refresher course.*
- Counting a teacher inservice day to take the test and district-sponsored workshops, the total public cost was $35.5 million. (*Alternative uses of these dollars to serve the same end might have been to create a fund to support the legal costs of districts seeking to fire incompetent teachers.*)
- Private costs in teacher time and preparation expenditures were an additional $42 million. (*Alternative uses of this resource might have been to require more advanced study by teachers..*)

The authors' clear preference was to preserve the status quo, avoid accountability requirements, and continue the citizens' reliance on input measures and trust in the education schools' quality control to provide the teachers who taught their children at their expense. The authors also criticized the TECAT as simplistic, too narrow in format, and too general in content, but they didn't advocate a "better" testing program. They favored eliminating teacher tests.

Perhaps the most ironic of the authors' opinions coupled two conflicting assertions — that the test was easy, simplistic, and beneath the dignity of professional educators, and so studying for the test should not be counted as a benefit. But, at the same time, the teachers, their union, and the school districts were afraid that many would fail the test, so a massive effort was undertaken to prepare the teachers for it and that should be counted as a cost.

Their point-of-view colored their analysis and affected their benefit-cost calculations.

Table 1 lists the costs and benefits as the authors calculated them. The authors decided that the TECAT produced a $53 million net cost. In their view, the development of the test itself accounted for only about $5 million, or less than 10 percent of the total cost of the test of $54 million. Three general categories of costs comprised the rest of the total: a day's worth of teachers' inservice time used for the test administration; time and materials used in test preparation workshops organized by the school districts; and time and materials spent by teachers privately.

Table 1: Costs and Benefit of the Texas Teacher Test According to Shepard, Kreitzer, and Graue (figures are supposed to represent annualized values)

Authors' estimate	Cost component
-$5,065,500	Test development & administration
-26,260,000	Teacher inservice day to take test
- 4,150,000	Preparation workshops and review - done by the school districts and, thus, paid for by the taxpayer
-43,152,000	Private teacher costs in study time, paid-for workshops, and purchased study materials
+25,295,466	Salary savings from dismissed teachers
-$53,470,534	Authors' TOTAL

The authors' benefit-cost analysis of the TECAT, however, is full of mistakes. The mistakes take several forms:

- arbitrary inclusions or exclusions of benefits or costs;
- miscalculations of the value of time, specifically, the value of teachers' after-hours time and the compounded value of recurring benefits
- counting gross costs when net costs (that include the value of countervailing benefits) would be more appropriate and the authors are making net cost conclusions;
- counting average costs when marginal (or, incremental) costs would be more appropriate and the authors are making marginal cost conclusions;

Correcting just for the more obvious of their mistakes, pushes the TECAT program's net benefits into the black, and by a wide margin. The authors' -$53 million turns into +$300 million. A summary listing of the authors' mistakes is provided in Table 2. A summary of the recalculation is provided in Table 3.

Here's how the recalculation is done. Taking their mistakes one-at-a-time, and in order:

Arbitrary Inclusions or Exclusions.

Cost of One In-Service Day. The authors arbitrarily include as a cost at least one item that should not be counted as a cost — $26 million for the teacher in-service day used for the test administration. By counting this as a pure cost, the authors assumed that the best alternative use for the teachers time would have been something they considered to be fully worthwhile, like a day's teaching, or in-service workshops with topics that each teacher chose herself to meet her own most significant needs (no comment on whether teachers generally spend their in-service time productively). This implies, however, that if it had not been for the TECAT there either would not have been an in-service day or the teachers would have been able to choose their in-service topic.

Such was not the case. In Texas at the time, half of the required in-service days' topics were of the teachers' choosing, but the other half were reserved by the education authorities for subject matter of particular general import. On these latter days, the state or district education authorities required that teachers participate in in-service activities of the authorities' choosing. On the test administration day in 1986, the teachers were scheduled to participate in an in-service activity of the authorities' choosing. (Texas Education Agency, 1994) The authorities decided that teacher literacy was the most pressing issue at the time and determined that that particular in-service day would be devoted to that tissue, as was their prerogative. The test format was their chosen means for inducing the teachers to study their literacy skills if they needed to. The "next best use" (i.e. the "opportunity cost") of the in-service day was for some other topic *necessarily* of less importance (necessarily, because if another topic had been more important, it would have been chosen).

The authors cannot legitimately count the teachers' time as a cost, then, unless they wish to argue that taking a high-stakes test (and the studying and heightened attention that it induces) is an inferior form of learning to sitting passively in a chair while a lecturer talks at you. In order to justify counting the in-service day a total loss (a pure cost) as they do, they must have concluded that one learns nothing from tests or the process of studying for and taking tests and that one always learns a significant, appreciable amount by attending someone else's lecture.

Presumably, the authors would have considered that a *lecture* on literacy skills would have been an acceptable, "non-costly" use of the teachers time on that in-service day.[1]

Benefit of Dismissing Illiterate Teachers. The only benefit included in the entire study was calculated as the sum of the salaries of the teachers who were dismissed after failing the TECAT after multiple opportunities to pass it. If Texans' could hire a literate teacher to replace an illiterate one, they were presumably getting value for their investment that they previously had not.

This may seem like a crude way to count the benefits, but it's really not so bad. Teachers were given a lot of help to pass the TECAT — study sessions, coaching, and multiple chances. And, by all accounts, the test was extremely easy. It was highly likely that if a teacher couldn't pass the TECAT, the teacher *was* illiterate. New teachers hired to replace the dismissed ones would have had to pass the TECAT, and so were far more likely to be literate.

But the authors decided that only some of the dismissed teachers were relevant to this issue, counting the salaries of 887 of the dismissed teachers, while not counting the salaries of the other 1063. Their rational for this exclusion was that the 887 counted were in "academic" jobs where their illiteracy could adversely affect the quality of their teaching, while the 1063 not counted were in "non-academic" jobs where their illiteracy allegedly would not affect the quality of their work.

Superficially, the principal makes sense. A custodian, for example, may not need basic literacy skills to perform her job effectively, whereas teachers obviously do.

So, which employees did the authors' categorize as "non-academic?" The answer: Kindergarten teachers; Music and Art teachers; ESL teachers; Industrial Arts teachers; Business education teachers; Counselors; and Physical Education teachers.

The authors asserted that, though literacy skills affect the quality of teaching in "academic" subjects, they are not important to these subjects. "Non-academic" teachers and students would likely feel shocked and insulted to hear this assertion. Parents of the affected students would likely feel outrage. The authors' assertion seems elitist and economic-class biased. It was also very presumptuous. The state of Texas decided that minimal literacy skills should be required of Kindergarten, Vocational Education, and the other groups of teachers the authors wished to exclude. And it was, after all, Texas' decision to make, not the authors'.

Including *all* the dismissed teachers as the authors should have done, one calculates an annual benefit of $55,610,100 million rather than the $25,295,466 million of the authors. The authors arbitrarily excluded some pertinent benefits.

The authors also refused to consider another 8,000 teachers who never showed up to take the TECAT. While it is certainly fair to assume that most of these were probably teachers who had planned to leave the teaching profession, or leave Texas, anyway, probably some number of them were teachers who decided, while studying for the TECAT, that they couldn't pass it.

Miscalculating the Value of Time.

Moreover, benefits from the dismissal of illiterate teachers recurred; they continued for years afterwards; they were not just one-time-only benefits.[2] A dismissed teacher was prevented from teaching for years[3], one should presume for the average remaining duration of the average teacher's career. Applying a discount rate of 7% and assuming conservatively that the illiterate teachers would

---

[1] I'm avoiding slipping into a more general topic here -- that is, are teacher in-service days, in general, used productively. Some would argue that they are not. The requirements on the part of teachers often amount to just showing up for a lecture or a series of lectures.

[2] James Caterall noticed this, too. See page 4 of his report.

[3] Though dismissed teachers could keep trying, as often as they wished, to pass the TECAT, which was administered a few times each year.

BEST COPY AVAILABLE

beside the point, too, because it's clear that the citizens of Texas wanted some accountability in their teacher certification system, and would not have been content with the minor modifications of the status quo — consisting of more input requirements — that the authors recommend.

Even the authors admitted that *half* the teachers interviewed thought that the test accomplished its purpose, "to weed out incompetent teachers and reassure the public."

Table 2 summarizes the mistakes in the authors' analysis.

Table 3 recalculates the authors' base numbers, correcting for mistakes in their calculations.

Table 2: *The Texas Teacher Test* as a Benefit-Cost Analysis — A Summary of the Cost Mis-Allocations

- arbitrary inclusions or exclusions of benefits or costs:
  - -- exclusion: more than half the dismissed teachers "don't count" in the benefit calculations -- those in vocational education, industrial arts, special education, business education, and kindergarten -- because, the authors argue, literacy is not important in their work.
  - -- inclusion: teacher time spent taking the test during one of their prescribed-topic in-service days is counted as a pure cost, implying that tests are not acceptable vehicles for teaching subject matter, while passive accumulation of seat time during lectures (with no accountability for listening) is an acceptable vehicle for learning
- miscalculations of the value of time:
  - -- they value teachers' after-hours time at their full salary rate; and
  - -- they ignore the future value of recurring benefits
- counting gross costs when net costs (that include the value of countervailing benefits) would be more appropriate and the authors are making net-cost conclusions;
  - -- (eg. Why would teachers take the nonrequired test preparation workshops on their own time if there was no benefit to be had?)

Correcting just for the more obvious of the authors' mistakes, pushes the TECAT program's net benefits into the black, and by a wide margin. The authors' -$53 million turns into +$330 million. This leaves the purchase costs of the exam (the "nominal" cost) as the only valid net cost item.

Table 3: Social Costs of the Texas Teacher Test According to Shepard, Kreitzer, and Graue —Corrected for Mistakes

Amount	description/correction
-$53,470,534	Authors' TOTAL
+30,314,634	correcting for the arbitrary exclusion of voc-ed, special ed. and other "non-academic" teachers as unneedful of basic literacy skills
-23,155,900	Sub-total
+283,625,416	accounting for the recurrent nature of the benefits
+260,469,516	Sub-total
+26,260,000	correcting for the arbitrary inclusion of the teachers' prescribed-topic in-service day that was used for the test administration as a pure cost
+286,729,516	Sub-total
+22,026,000	correcting for the overvaluation of teachers' personal time and the mystery of the $30/hour workshops
+308,755,516	Sub-total
+23,826,000	accounting for the countervailing benefits of the teacher workshop and study time
+332,581,516	CORRECTED TOTAL

Other Benefits.

Tests provide information. And they do so rather inexpensively. It turns out that the teachers TECAT score, as simple a measure as it may have been, was correlated with student achievement. Ron Ferguson used TECAT scores and other data in an effort to predict student achievement in Texas. (see Ferguson, 1991a and 1991b) His articles have been hailed as some of the first to show evidence that higher levels of education spending can, in certain circumstances, produce higher levels of student achievement, in this case because teachers with higher TECAT scores tended to get paid more.

While controlling for other, traditionally-used predictors, he found that the TECAT score provides additional, significant predictive power. In the lower grades, a teacher's TECAT score was the strongest single predictor of student achievement — the same test that Shepard, et.al. described as meaningless and worthless. Ferguson found that the teacher's TECAT score explained 48 percent of the variance in predicting 5th-grade student reading scores in majority black school districts, and 26 percent in predicting 9th-grade student reading scores in majority black school districts. The influence of teacher TECAT score was stronger than that of any other predictors, including parents' education, family poverty level or number of parents, or class and district size. The ethnicity of the teacher was not a significant predictor. Teacher TECAT score was the strongest predictor for 5th grade reading scores in majority hispanic school districts, too, and the second strongest predictor of 9th-grade reading scores in majority hispanic school districts.

He ended his studies with quite the opposite attitude of the authors of the TECAT case study. He noticed that disadvantaged and minority children tended to get the teachers scoring lowest on the TECAT, and he thinks it was unfair...to the students.

Private Net Benefits — Texas Teachers.

The authors strongly imply that the TECAT was a failure not only for all Texans, but for the teachers and other educators, especially. The TECAT embarrassed the profession because the teachers made such a fuss over an exceptionally easy exam, and some failed. The TECAT caused unnecessary stress, and the authors view stress as a cost. The message of the authors is, don't make such a deal again.

But a brief look at the benefits and costs suggests a different story. The teachers who didn't lose their jobs won big (see Table 4). Table 4 counts the salary increase that the teachers won as part of the

TECAT deal, along with a promise of smaller class sizes (thus, a reduced workload) and an increase in the length of the kindergarten day from half- to full-day (thus, more work and more pay for some teachers and greater seniority for others). (Incidentally, these costs did not get counted in either the authors' social accounting or mine, for legitimate reasons. These amounts are "transfers" from the taxpayers to the teachers. From society's perspective, the cost to the taxpayers is equally balanced by the benefit to the teachers and, thus, there is no net cost or net benefit to society.)

Table 4: Private Benefit-Cost Analysis of the Texas Teacher Test Using Authors' Base Numbers, but Recalculated to Reflect Texas Teachers' Perspective, with Salary Increase Included

(all figures are net present values, single annual figures)

Amount	Description
-$43,766,000	Private teacher cost of workshops, materials, and supplies
+$1,732 ,100,000	Salary increase over five years (one year= $394.9 million — my estimate using figures from Shepard report)
(+)	Smaller class size (i.e. reduced workload)
(+)	1/2 day kindergarten required, adding more jobs
+ $1,732,100,000 + reduced workloads + more jobs	CORRECTED TOTAL (annualized, for five year period)

What was a benefit to the teachers, however, was a cost to the taxpayers. Why were they willing to pay so much? The authors made it clear that the legislative deal was a straight *quid pro quo* — the teachers got their salary increase, reduced workloads, and full-day kindergarten and the taxpayers got the TECAT — some assurance that their teachers met a minimal level of literacy. How much was that assurance worth to Texas taxpayers? Quite a lot. Table 5 summarizes Texas' willingness-to-pay for an assurance that their teachers were minimally literate. "Willingness-to-pay" is simply empirical evidence of the level of demand.

Texas' willingness-to-pay for an assurance of teacher literacy also demonstrates that taxpayers *are* willing to pay more for education *if they can get something in return* from the education system.

Table 5: Texas' "Willingness-to-Pay" for Assurance of a Minimal Level of Teacher Literacy

· *expected* nominal cost of test ($3 million)
· teachers' salary increase ($1,723 million over 5 years)
· salaries of new teachers hired
- to reduce class sizes
- to extend the kindergarten day

TOTAL? Well over $1,726 million, or 4 percent of education budget annually.

BEST COPY AVAILABLE

Discussion.

If the authors were right about public opinion, the TECAT controversy was a messy affair that left many people disappointed. Furthermore, according to the authors at least, the TECAT produced unanticipated costs. So, the authors concluded, don't have teacher tests.

What of the concerns of the citizens of Texas that there was no accountability in the teacher certification system? The authors recommended putting more money into the same system, perhaps changing some input requirements, and making no changes to the status quo.

An alternative that the authors didn't consider would have been to move the TECAT to an earlier point in the teacher training process, say at the end of graduate school, or even at the beginning of graduate school. This would have met the concerns of the citizens of Texas. It would have achieved all the same benefits. But, most of the costs that the authors enumerated would have evaporated. There would have been no loss of teacher time. The responsibility for preparing the teachers for the test would have been placed on the teacher training schools or, better, on the potential education students themselves. And, best of all, the time of unqualified would-be teachers (and their students) would not have been wasted.

A reasonable alternative to the authors' complaints about the alleged simplistic nature of the TECAT would have been to initiate a required "higher-level" exam for teachers, in addition to the TECAT.

As it turns out, the citizens of Texas did not follow the authors' advice. Rather, they followed the path just drawn, making the basic literacy exam an entrance exam for education school and requiring new teachers to pass another, newly-created exam that focussed on each teacher's content area and on pedagogy and professional development. They increased the benefits and reduced the costs, even according to the authors' benefit-cost accounting criteria. And they ended up with more tests, not fewer.

In another, separate article, economists Lewis Solmon and Cheryl Fagnano conducted a more sophisticated analysis of the data from the Shepard, Kreitzer, and Grau study and found that net benefits could be in excess of $1,250 million. Among other things, they estimated the value over many students' lifetimes of the increased learning they would gain from more literate teachers. (see Solmon and Fagnano, 1990)

The Fractured Marketplace for Standardized Testing

A few years ago, while working at the General Accounting Office, I was assigned the task of estimating the cost of a national examination system, a concept then very much in the news. In my attempt to learn the unit costs of the types of tests proposed for national exams, I was surprised by the paucity of information available on the subject. I found studies here and there based on small samples of school districts and some incomplete cost information derived from larger samples. However, no study appeared to exist that provided complete cost information based on micro-data. Test publishers with relatively good information were not willing to reveal it for fear of informing their competitors.

A study did exist, however, that used *macro-data* (i.e. nationally aggregated totals) to estimate testing costs. I learned that the Ford Foundation funded a group calling itself the National Commission on Testing and Public Policy to do the job. The group boasted a few well-known members, including the then-governor of Arkansas, Bill Clinton.

In 1990, with staff provided by the Test Policy center at Boston College's education school, the group issued a report entitled *From Gatekeeper to Gateway: Transforming Testing in America*. That report criticized standardized testing and claimed emphatically that such testing was overly costly in terms of both money and time. U.S. students, it asserted, "are subjected to too much standardized testing" and standardized testing "devours" teaching time and "looms ominously" in students' lives. The report went on to claim that "mandatory testing consumes some 20 million school days annually." (Twenty million is admittedly a large number; but there are a large number of U.S. students — about 40 million. The average, which the report did not compute, is one-half day of testing per student per year.)

Alluding to information on the extent and cost of testing, the report suggested that comprehensive data was available on the subject, and could be found in background sources. The footnotes to the 1990 report, however, referred only to a book "in press." That book is the one under

review here, *The Fractured Marketplace for Standardized Testing*, from the Boston College Test Policy center, which arrived at the press room in early 1993. At one point in the book, the authors, Walter Haney, George Madaus, and Lyons refer to an earlier testing program evaluation that they used as a guide in designing their own study. It was *A Case Study of the Texas Teacher Test*.

Fractured Marketplace's Methods.

With most books, following a footnote to its source is not very exciting or surprising. Not so with *Fractured Marketplace*. Take, for example, footnote 11 on p. 108 of *Fractured Marketplace*. It follows this sentence in the text: "As a plausible figure, we assume that for every test battery, teachers and students devote 20 hours of classroom time to test preparation activities." This statement is the most important in their entire book, because this "plausible" figure ends up accounting for 78% of the authors' estimate of the cost of standardized testing programs.

Leaving aside for the moment the question of the value of the time spent preparing for standardized tests (the authors think it has zero value), let's look at the authors' 20-hours-of-classroom-time number. First, there is no empirical basis for the number. Though it certainly is plausible under certain circumstances, it is far from being a U.S. average that could validly be used as a universal multiplier to estimate the cost of all standardized testing in the United States, which is how the authors use it. The best evidence shows average test preparation time to be about equal to average test-taking time (rather than 5 times test-taking time, as the authors assert), placing the authors' claim about 41 standard deviations above the average. (see U.S. General Accounting Office, 1993, pp. 18–19).

Second, let's follow the authors' footnote. It refers us to a particular section in the Congressional Office of Technology Assessment's report *Testing in American Schools: Asking the Right Questions*, which profiles the costs of testing in a single U.S. school district. Haney, Madaus, and Lyons use this single anecdotal example as the only support for their 5-times-test-taking-time multiplier — a single school district, out of over 12,000 in the United States. The OTA report reads: "In conversations with district teachers, OTA found that the time they spend in classroom *preparation* of students for the standardized tests varies from 0 to 3 weeks per testing administration." Haney, Madaus, and Lyons identify 1.5 weeks as the "median" amount of time used in test preparation.

But, the remainder of the same paragraph in the OTA report is very revealing. It reads, "Some teachers claim they spend no time doing test preparation that is distinguishable from their regular classroom instruction; others use the standardized test as a final examination and offer students the benefit of lengthy in-class review time." In other words, there is no difference between test preparation time and regular class time in this school district. Either the test has no effect on regular classroom instruction or it is totally integrated into the instruction, an inseparable and indistinguishable part of the curriculum. There goes 78 percent of Haney, Madaus, and Lyons' estimate for the cost of standardized testing.

Nominally, *The Fractured Marketplace* is full of quantitative detail, and all the data and calculations create an illusion of precision. But most of the calculations are built atop flimsy base numbers and carried out with untenable assumptions. One result is a incredibly wide range between the authors' "low" and high estimates for the cost of standardized testing. The authors "low" estimate for the cost per student-hour of state and local district tests is about $3.50, while their high estimate is about $66.

Double-Counting Costs.

Moreover, while the authors' high estimates are based on wishful thinking, even the "low" estimates are exaggerations. For example, 12 percent of their "low" cost estimate for state and district testing is overhead — the cost of operating a school facility during test and test-preparation times. It might be legitimate to count this as a cost of testing if space actually had to be rented for the occasion of a test administration.[6]

---

[6] But such is rarely the case, even with tests administered by private firms, such as the ACT, the SAT, and the AP examinations. With these exams, the firms pay no rental fee because schools

But, in the case of state and district tests, the tests are, in most cases, administered during the regular school year, during regular school hours. Overhead costs are already paid for (or, "sunk"); they are already paid for whether the school building is used for a test, used for anything else, or not used at all. Using a school building for a test incurs no *additional* overhead cost. But, the authors imply that it is an additional cost — a cost incurred *because* a test, rather than some other activity, occupies a certain time period. That's double-counting.[7]

Table 6 adjusts some of the authors' calculations for the cost of state and district testing. Even these adjustments do not address all of the authors' cost items but, as the reader can see, the corrected "high" estimate is already below their original lower bound.

Table 6: **Recalculated Costs of State and District Testing Programs**
as per authors' convention, all figures represent dollar cost/student/hour
state and district costs are combined into a single average

authors' "low" estimate	authors' high estimate	description/correction
$3.50	$66.00	authors' estimate of the total cost/student/hour
	-51.58	subtraction of test preparation time multiplier (5x test time)
	14.42	sub-total
-0.41	-2.50	double-counting of building rental
3.09	11.92	sub-total
-1.01	-9.01	double-counting student test-taking time[8]
2.08	2.91	sub-total
-0.54	-0.75	overestimate testing-time per student by 35%[9]
$1.54	$2.16	PARTIALLY CORRECTED TOTAL

Valuing Student Time.

Most of the cost of testing according to Haney, Madaus, and Lyons is accumulated by simply assuming, on their part, that student test-taking and test-preparation time have no value, instructional or otherwise. They are a pure loss to learning. In their view, students learn nothing while taking tests. And, they learn nothing while preparing for tests, which implies, of course, that their teachers are deliberately wasting the students' time in an activity that they should be able to discern (if the teachers have any power of observation) to have no instructional payoff. Presumably, students learn only while listening to teachers; they can't learn anything on their own.

---

give them space free of charge. The schools assume that many of their own students will be taking the exams and administering them on campus makes it more convenient for the students.

[7] An example of a situation where it would be appropriate to attribute *some* overhead costs to testing involves a test that is administered on a day when a school building would otherwise not even be open. Any custodial or security team costs incurred because of the exam administration and for no other reason could legitimately be counted as marginal overhead cost induced by an exam.

[8] One can accept student test-taking time as a cost only by believing a set of assumptions, any one of which seems far-fetched. These assumptions are addressed at length in the rest of this article, starting in the next section.

[9] See U.S. General Accounting Office, pp. 18–19 for a better estimate.

Tests do require activity on the students' part that differs from teacher lectures or other teacher-led classroom activities. In some ways, it is more like homework — students are given questions to answer or projects to do according to a set of instructions and they do the work on their own. One wonders if Haney, Madaus, and Lyons would argue that students learn nothing while doing homework.

If they would, that would mark a further indictment on the authors' part of teachers' and local administrators' abilities to discern when their students are learning and when they are not. It is obvious that most teachers and local administrators think that the benefits of tests outweigh the costs, up to a certain saturation point.[10] It is obvious because most teachers and local administrators use tests as part of their instructional programs.

The authors calculate the cost of the time devoted to testing by assuming that seat time in school is directly translated into a lifetime's earnings. The more time you spend in classrooms listening to teachers, the more you will earn in your lifetime. If you miss some of this time due to taking tests, you will be poorer.

Table 7 summarizes the assumptions that Haney, Madaus, and Lyons make in counting student test-taking time as a cost.

Table 7: Counting Student Test-Taking Time as a Cost

Assumptions:

- students learn nothing while taking standardized tests
- students learn nothing from the process of taking tests
- tests are unrelated to the other elements of instruction; there is no mutual dependence and no interaction effect

Counter-argumements to the author's valuation of student test-taking time:

- Students learn while studying for tests and while taking tests. Moreover, they learn from the process of taking tests and that's important. There are many tests in life — in job-hunting and in courtship, for example — so we would be doing our children no favors by hiding this inevitable aspect of life from them in school.
- Indeed, students may learn *more* through test-taking than through passive listening. Students may learn better because they are forced to actively use their mental faculties. They are forced to think, and think for themselves. That isn't necessarily the case during regular class time.
- The labor market pays attention to diplomas, and to measures of achievement if they exist, not seat time in school.
- Part of the reason reformers want to impose accountability requirements is that

---

10 The GAO study on the extent and cost of standardized testing asked its national sample of state and local school administrators if the benefits of testing outweighed the costs. Respondents from both groups strongly believed that the net benefits of their current testing programs (in 1990-91) were positive. Seventy-five percent of state respondents felt that way (compared to 5 percent who felt the opposite) and 43 percent of local respondents felt that way (compared to 18 percent who felt the opposite). State respondents believed strongly that net benefits would increase if their testing programs were somewhat larger — 52 percent indicated so (compared to 5 percent indicating the opposite). At the local level, slightly more respondents (28 percent versus 22 percent) though net benefits would decrease than thought they would increase with a somewhat larger district testing program, but 40 percent thought net benefits would remain the same. Thus, 62 percent of local respondents thought net benefits would increase or remain the same with an additional test.

they think too much time is now being wasted in non-instructional activities. They think that adding tests will *increase* instructional time. They may be right. Adding tests may also increase study time.

The Testing Industry.

The Fractured Marketplace is more than just a benefit-cost analysis, however. Another part of the book provides an encyclopedic history of standardized testing (with an emphasis on its shortcomings and the misuse of test results) and an overview of the standardized testing industry. This part of the book is informative, though the tone is incessantly negative. The authors retread decades of criticism of standardized testing (these are also "costs" of testing, broadly defined) and make not the faintest effort to enumerate the benefits. Indeed, one wonders where these fellows have been the past couple decades. Much has changed in the character of standardized testing and in its administration, but the authors seem not the slightest bit aware of it.

Probably the single most important recent innovation to improve the quality and fairness of testing in the United States is the addition of managerial and technical expertise in state education agencies. At that level, it is possible to retain an adequately-sized group of well-paid, technically-proficient testing experts, adept at screening, evaluating, administering, and interpreting tests. These people are not beholden to test publishers. They are not naive about test results. And, they, along with governors and legislatures, are currently calling the shots in standardized testing. Some of the most important decisions that affect the design and content of standardized tests, and the character of the testing industry and the nature of its work, are today being made by state testing directors.

In the same section of *Fractured Marketplace*, on page 54, the term "fractured marketplace" is defined:

> "Different firms are engaged in different segments of the testing marketplace. And even for a single test, different organizations may be responsible for sponsoring, building, administering, scoring and reporting that test. Also, while there have been a significant number of mergers and acquisitions among firms active in the testing marketplace over the last twenty years, it is clear from the new prominence of firms such as Scantron and PRO-ED that the testing industry remains fluid enough to allow the successful entry of new players. And the rapid rise to eminence of firms such as NCS and Scantron shows that computer technology is having an increasing influence on the testing marketplace and that test-related services, such as scoring and reporting of results, are an increasingly important segment of the market, as compared with sales of tests themselves."

The passage describes a market with many players, overlapping niches with no monopoly positions, and an ease of entry and exit. In other words, a "fractured" market is an ordinary, healthy, competitive market —the kind that serves the consumer best. Why, then, is the final chapter in the book devoted to "Mending the Fractured Marketplace?" It is only reasonable to conclude that it is because the authors favor centralized bureaucratic control over student testing, managed by overseers such as themselves or others with similar ideological leanings.

Discussion

Concerned and well-trained educators in state agencies and local school districts make the decisions to purchase or develop standardized tests. They are neither as unfair, nor as sinister, nor as oblivious of the limitations of those tests as the authors would make them out to be. Simply put, educators purchase or develop tests because they believe that the benefits outweigh the costs. Furthermore, the use of standardized testing has increased over the years because the technology has gotten better and more efficient, the tests are employed more fairly now, and several needs are better met through standardized testing than through alternatives.

The authors have two recommendations to preserve education from the alleged disaster of standardized tests. One is to reduce our supposed over-reliance on them through a "refusal to accept bondage to a single technology." It may be a fair point that students should not be judged summarily on scores from a single test taken at a single point in time. It must be added, however, that few U.S. school districts make such a single-test judgement.

BEST COPY AVAILABLE

Without standardized tests we would be left with only one other "technology" for evaluating student or school performance — grade point averages. Should we as a society (to paraphrase Haney and Madaus) "accept bondage to that single technology?" Particularly in the United States, with its lack of uniform academic standards, schools vary widely in quality and in the rigor (indeed, in the *philosophy*)[11] of their grading practices. A 3.0 gpa at school A is not necessarily equivalent to a 3.0 gpa at school B. Grade point averages are *norm-referenced* measures of performance, normed at the *school level*!

But, if they had their way, this is all that Haney, Madaus, and Lyons would let us use. There is another cost, of course, to a reliance on such poor and limited information, when much more information (from standardized tests) could be used. But, the authors don't calculate the magnitude of that cost.

The authors' other recommendation involves the plan a couple of years ago to form a national commission that would evaluate tests proposed for inclusion in an eventual national examination system. The authors object to the plan because the commission members would be appointed by our elected representatives — the President, members of both parties in Congress, the governors, and state legislators. Instead, they propose a "truly independent" national commission with academics like themselves.

Education research, as a collective enterprise, has a problem. Much of it is colored ideologically or by interest-group bias. The two works discussed here may be cases in point. I say that not because they have errors, but because all the errors point in the same direction — toward exaggerating the cost of testing and minimizing the benefits.

Standardized tests, after all, can be used for more than simply measuring individual students' performance. They can be, and are, used to judge the performance of the education system and, as such, are a threat to the interests vested in that system. My home-town newspapers already glibly dismiss education research that is critical of testing as part of a "kill the messenger mentality." These two studies could do nothing but further that skeptical perspective.

Public support for a greater use of standardized tests remains strong, however. (See Elam and Johnson) And, since the education system is *supposed to* serve the interests of the taxpayers who pay the bills and the parents who entrust their loved ones to it, perhaps it should administer more tests, not fewer.

REFERENCES

Council of Economic Advisors, (1989). *Economic Report of the President*. Washington, D.C.: U.S. Government Printing Office.

Elam, Stanley. *How America Views its Schools*. Bloomington, IN: Phi Delta Kappa, 1995.

Ferguson, Ronald F. (1991a). "Paying for Public Education: New Evidence on How and Why Money Matters," *Harvard Journal on Legislation, 28(2), 465-498.*

Ferguson, Ronald F. (1991b). "Racial Patterns in How School and Teacher Quality Affect Achievement and Earnings," paper presented at the annual meeting of the American Association for Public Policy Analysis and Management, Bethesda, Md.

Friedman, Lee S. (1984). *Microeconomic Policy Analysis*. New York: McGraw-Hill.

Gramlich, Edward M. (1990). *A Guide to Benefit-Cost Analysis*. Englewood Cliffs, N.J.: Prentice-Hall.

Haney, Walter M., Madaus, George F., and Lyons, Robert (1993). *The Fractured Marketplace for Standardized Testing*. Boston: Kluwer Academic.

Johnson, Jean and John Immerwahr, "First Things First: What Americans Expect from the Public Schools." New York: Public Agenda Foundation, 1994.

---

[11] If Rita Kramer's sensational stories are true, some education schools are encouraging teachers to grade students not according to performance but according to the teacher's perception of each student's socio-economic background. That is, students from minority groups and students whom a teachers judges to be poor would be graded easier, regardless of effort or performance.

Kramer, Rita. (1991). *Ed School Follies*. New York: Free Press.

Shepard, Lorrie A., Kreitzer, Amelia E., and Graue, M. Elizabeth (1987a). " A Case Study of the Texas Teacher Test: Technical Report "CSE Report No. 276. Los Angeles: National Center for Research and Evaluation, Standards, and Student Testing, University of California.

Shepard, Lorrie A. and Kreitzer, Amelia E. (1987b). "The Texas Teacher Test, " *Education Researcher, 16(6), 22-31.*

Solmon, Lewis C. and Cheryl L. Fagnano, "Speculations on the Benefits of Large-scale Teacher Assessment Programs: How 78 Million Dollars Can be Considered a Mere Pittance," *Journal of Education Finance*, Vol.16, Summer, 1990, pp. 21–36.

Stokey, Edith and Zeckhauser, Richard. (1978). *A Primer for Policy Analysis*. New York: Norton.

Texas Education Agency, Division of Assessment, conversations with Rick Bataille and Nolan Woods, February through May, 1994.

U.S. Congress. Office of Technology Assessment. (1992). *Testing in American Schools: Asking the Right Questions*. OTA-SET-519. Washington, D.C.: U.S. Government Printing Office.

U.S. General Accounting Office. (1993). *Student Testing: Current Extent and Expenditures, With Cost Estimates for a National Examination*. GAO/PEMD-93-8. Washington: author.

Richard P. Phelps is a research analyst at American Institutes for Research, Pelavin Research Center, Washington, D.C.